Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 97 (2006) 1984 – 1996
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

James–Stein estimators for time series regression models ☆

Motohiro Senda, Masanobu Taniguchi*

Department of Mathematical Sciences, School of Science and Engineering, Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

Received 12 March 2005
Available online 11 July 2006

Abstract

The least squares (LS) estimator seems the natural estimator of the coefficients of a Gaussian linear regression model. However, if the dimension of the vector of coefficients is greater than 2 and the residuals are independent and identically distributed, this conventional estimator is not admissible. James and Stein [Estimation with quadratic loss, Proceedings of the Fourth Berkely Symposium vol. 1, 1961, pp. 361–379] proposed a shrinkage estimator (James–Stein estimator) which improves the least squares estimator with respect to the mean squares error loss function. In this paper, we investigate the mean squares error of the James–Stein (JS) estimator for the regression coefficients when the residuals are generated from a Gaussian stationary process. Then, sufficient conditions for the JS to improve the LS are given. It is important to know the influence of the dependence on the JS. Also numerical studies illuminate some interesting features of the improvement. The results have potential applications to economics, engineering, and natural sciences.

AMS 1991 subject classification: Primary 62J07; 62M10; Secondary 62C15; 62E20

Keywords: James–Stein estimator; Least squares estimator; Gaussian stationary process; Mean squares error; Time series regression model; Regression spectrum; Residual spectral density matrix

1. Introduction

Let $\mathbf{y}(1), \ldots, \mathbf{y}(n)$ be a sequence of independent and identically distributed random vectors distributed as the k-dimensional normal distribution $N(\boldsymbol{\theta}, \mathbf{I}_k)$, where $\mathbf{I}_k$ is the $k \times k$ identity matrix. As an estimator for the population mean $\boldsymbol{\theta}$, the sample mean $\overline{\mathbf{Y}}_n \equiv n^{-1} \sum_{t=1}^{n} \mathbf{y}(t)$

☆ This work was supported by Waseda University Grant for Special Research Project 2004 A-163.

* Corresponding author.
E-mail addresses: motohiro_senda@moegi.waseda.jp (M. Senda), taniguchi@waseda.jp (M. Taniguchi).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2005.08.011

seems the most fundamental and natural estimator. However, if $k \geqslant 3$, Stein [8] showed that this conventional estimator is not admissible with respect to the mean squares error (MSE) loss function. Furthermore James and Stein [7] propose a shrinkage estimator (JS estimator) $\widehat{\boldsymbol{\theta}}_n \equiv \{1-(k-2)/\|\overline{\mathbf{Y}}_n\|^2\}\overline{\mathbf{Y}}_n$, which improves $\overline{\mathbf{Y}}_n$ with respect to MSE when $k \geqslant 3$. Here $\|(\cdot)\|$ is the Euclidean norm of $(\cdot)$. The analysis has been extended to the case of estimation for the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$. Similarly, based on the least squares estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ of $\boldsymbol{\beta}$, estimators of JS type have been proposed. In the case where $\boldsymbol{\varepsilon}$ is distributed as $N(\mathbf{0}, \sigma^2\mathbf{I}_n)$, Arnold [2] proposed an estimator $\widetilde{\boldsymbol{\beta}} \equiv \{1 - c\widehat{\sigma}^2/\|\mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}\|^2\}\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$, where c is a positive constant number and $\widehat{\sigma}^2$ is an estimator of σ^2. This estimator improves $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ with a suitable choice of c if $\dim \boldsymbol{\beta} \geqslant 3$.

We often use statistical methods designed for independent sample even for the case when the observations might be dependent (e.g., financial engineering analysis). In such a case, it is very important to elucidate the influence of dependence on the statistical methods. In this paper, we investigate the statistical properties for estimators of JS type in time series regression models. For time series data, estimation of the trend is very important in various fields, e.g., econometrics, engineering and natural sciences. Usually, we estimate the coefficients of regression part by the LS estimator. However, if the regressors and the residual spectra satisfy some relations, the LS becomes inferior to the JS. For a typical time series model with cyclical trend, this point is described in Example 2 clearly. In this case, the JS estimator becomes an alternative one.

For a vector-valued Gaussian stationary process with mean vector $\boldsymbol{\mu}$, Taniguchi and Hirukawa [9] studied the MSE of sample mean $\widehat{\boldsymbol{\mu}}$ and the JS estimator $\widehat{\boldsymbol{\mu}}_{\mathrm{JS}}$ for $\boldsymbol{\mu}$. Then they gave a set of sufficient conditions for $\widehat{\boldsymbol{\mu}}_{\mathrm{JS}}$ to improve $\widehat{\boldsymbol{\mu}}$ in terms of the spectral density matrix. In this paper, letting $\mathbf{B}$ be a matrix of unknown coefficients, we consider the following time series regression model

$$\mathbf{y}(t) = \mathbf{B}'\mathbf{x}(t) + \boldsymbol{\varepsilon}(t),$$

when $\{\mathbf{x}(t)\}$ satisfies Grenander's condition, and $\{\boldsymbol{\varepsilon}(t)\}$ is a vector-valued Gaussian stationary process with spectral density matrix $\mathbf{f}(\lambda)$. Then, a set of sufficient conditions for the JS estimator $\widehat{\mathbf{B}}_{\mathrm{JS}}$ of $\mathbf{B}$ to improve the LS estimator $\widehat{\mathbf{B}}_{\mathrm{LS}}$ is given in terms of $\mathbf{f}(\lambda)$ and the regression spectrum matrix. Numerical studies are provided, and they illuminate some interesting features of $\widehat{\mathbf{B}}_{\mathrm{JS}}$.

This paper organized as follows. Section 2 describes the model settings and assumptions, and compare the MSE of $\widehat{\mathbf{B}}_{\mathrm{JS}}$ with that of $\widehat{\mathbf{B}}_{\mathrm{LS}}$. Then, we give a set of sufficient conditions for $\widehat{\mathbf{B}}_{\mathrm{JS}}$ to improve $\widehat{\mathbf{B}}_{\mathrm{LS}}$ in terms of $\mathbf{f}(\lambda)$ and the regression spectrum. Here it should be noted that the MSE feature of $\widehat{\mathbf{B}}_{\mathrm{JS}}$ is greatly different between the case of $\mathbf{B} = \mathbf{0}$ and $\mathbf{B} \neq \mathbf{0}$. Hence we state the results in Theorems 1 and 2 separately. Numerical studies are given for various choices of $\mathbf{x}(t)$ and $\mathbf{f}(\lambda)$, and they illuminate the improvement of $\widehat{\mathbf{B}}_{\mathrm{JS}}$ to $\widehat{\mathbf{B}}_{\mathrm{LS}}$ visually and clearly. The proofs of theorems are relegated to Section 3.

There has been a series of works which deal with a spectral approximation of the BLUE estimator for $\mathbf{B}$ (e.g., [4]). It is known that it improves the LS uniformly, but, as we said, the motivation of this paper is not this point, i.e., we elucidate the influence of dependence on the JS estimator.

2. James–Stein estimators for time series regression models

Suppose that $\{\boldsymbol{\varepsilon}(t)\}$ is a p-dimensional Gaussian stationary process with mean vector $E[\boldsymbol{\varepsilon}(t)] = \mathbf{0}$ and spectral density matrix $\mathbf{f}(\lambda)$. In this paper, we consider the linear regression model of the form

$$\mathbf{y}(t) = \mathbf{B}'\mathbf{x}(t) + \boldsymbol{\varepsilon}(t), \qquad t = 1, \ldots, n, \tag{1}$$

where $\mathbf{y}(t) \equiv (y_1(t), \ldots, y_p(t))'$, and $\mathbf{B} \equiv (b_{ij})$ is a $q \times p$ matrix of unknown coefficients. Here $\mathbf{x}(t) \equiv (x_1(t), \ldots, x_q(t))'$ is a $q \times 1$ vector of regressors satisfying the assumptions below. Let $d_i^2(n) \equiv \sum_{t=1}^n \{x_i(t)\}^2$, $i = 1, \ldots, q$.

Assumption 1. $$\lim_{n\to\infty} d_i^2(n) = \infty, \quad i = 1, \ldots, q.$$

Assumption 2. $$\lim_{n\to\infty} \frac{\{x_i(n)\}^2}{d_i^2(n)} = 0, \quad i = 1, \ldots, q,$$

and for some $\kappa_i > 0$,

$$\frac{x_i(n)}{n^{\kappa_i}} = O(1), \quad i = 1, \ldots, q.$$

Assumption 3. For each $i, j = 1, \ldots, q$, there exists the limit

$$\rho_{ij}(h) \equiv \lim_{n\to\infty} \frac{\sum_{t=1}^n x_i(t)x_j(t+h)}{d_i(n)d_j(n)}.$$

Let $\mathbf{R}(h) \equiv (\rho_{ij}(h))$.

Assumption 4. $\mathbf{R}(0)$ is nonsingular.

From Assumptions 1–4, we can write

$$\mathbf{R}(h) = \int_{-\pi}^{\pi} e^{ih\lambda}\mathbf{M}(d\lambda),$$

where $\mathbf{M}(\lambda)$ is a matrix function whose increments are Hermitian nonnegative, and which is uniquely defined if it is required to be continuous from the right and null at $-\pi$. Assumptions 1–4 are a slight modification of Grenander's condition. Now we make an additional assumption.

Assumption 5. $\{\boldsymbol{\varepsilon}(t)\}$ has absolutely continuous spectrum which is piecewise continuous with no discontinuities at the jumps of $\mathbf{M}(\lambda)$.

We rewrite (1) in the tensor notation

$$\mathbf{y} = (\mathbf{I}_p \otimes \mathbf{X})\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{U}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

wherein $y_i(t)$ is in row $(i-1)p + t$ of $\mathbf{y}$, $\varepsilon_i(t)$ is in row $(i-1)p + t$ of $\boldsymbol{\varepsilon}$, $\mathbf{X}$ has $x_j(t)$ in row t column j, $\boldsymbol{\beta}$ has b_{ij} in row $(j-1)q + i$ and $\mathbf{U} \equiv \mathbf{I}_p \otimes \mathbf{X}$.

If we are interested in estimation of $\boldsymbol{\beta}$ based on $\mathbf{y}$, the most fundamental candidate is the least squares estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} \equiv (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{y}$. When $\boldsymbol{\varepsilon}(t)$'s are i.i.d., it is known that $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ is not admissible if $pq \geqslant 3$. Stein [8] proposed an alternative estimator

$$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}} \equiv \left(1 - \frac{c}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b})\|^2}\right)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b}) + \mathbf{b}, \tag{2}$$

which is now called the James–Stein estimator for $\boldsymbol{\beta}$. Here $c > 0$ and $\mathbf{D}_n \equiv \mathrm{diag}(d_1(n), \ldots, d_q(n))$. $\mathbf{b}$ is a preassigned $pq \times 1$ vector toward which we shrink $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$. Stein showed that the risk of $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ is everywhere smaller than that of $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ under the MSE loss function

$$\mathrm{MSE}_n(\widehat{\boldsymbol{\beta}}) \equiv E\left[\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2\right]$$

if $pq \geqslant 3$ and $c = pq - 2$.

In this section, we compare the risk of $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ with that of $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ when $\{\boldsymbol{\varepsilon}(t)\}$ is a Gaussian stationary process. First, from Gaussianity of $\{\boldsymbol{\varepsilon}(t)\}$, it is seen that

$$(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b}) \sim N((\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b}), \mathbf{C}_n), \tag{3}$$

where

$$\mathbf{C}_n \equiv (\mathbf{I}_p \otimes \mathbf{D}_n)(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathrm{cov}(\boldsymbol{\varepsilon})\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}(\mathbf{I}_p \otimes \mathbf{D}_n).$$

From Theorem 8 of Hannan [5, p. 216], if Assumption 5 holds, it follows that

$$\mathbf{C} \equiv \lim_{n\to\infty} \mathbf{C}_n = (\mathbf{I}_p \otimes \mathbf{R}(0))^{-1} \int_{-\pi}^{\pi} 2\pi \mathbf{f}(\lambda) \otimes \mathbf{M}'(d\lambda)(\mathbf{I}_p \otimes \mathbf{R}(0))^{-1}. \tag{4}$$

Let $v_{1,n}, \ldots, v_{pq,n}$ $(v_{1,n} \leqslant \cdots \leqslant v_{pq,n})$ and $v_1, \ldots, v_{pq}$ $(v_1 \leqslant \cdots \leqslant v_{pq})$ be the eigenvalues of $\mathbf{C}_n$ and $\mathbf{C}$, respectively.

We evaluate

$$\begin{aligned}\mathrm{DMSE}_n &\equiv E\left[\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \boldsymbol{\beta})\|^2\right] - E\left[\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{JS}} - \boldsymbol{\beta})\|^2\right] \\ &= -c^2 E\left[\frac{1}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b})\|^2}\right] \\ &\quad + 2c\left(1 - E\left[\frac{\langle(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b}), (\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b})\rangle}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b})\|^2}\right]\right).\end{aligned} \tag{5}$$

Because the behavior of DMSE_n in the case of $\boldsymbol{\beta} - \mathbf{b} = \mathbf{0}$ is greatly different from that in the case of $\boldsymbol{\beta} - \mathbf{b} \neq \mathbf{0}$, first, we give the result for $\boldsymbol{\beta} - \mathbf{b} = \mathbf{0}$.

Theorem 1. *In the case of* $\boldsymbol{\beta} - \mathbf{b} = \mathbf{0}$, *suppose that Assumptions* 1–5 *hold and that* $pq \geqslant 3$. *Then,*

(i)

$$\begin{aligned}&c\left\{2 - \frac{c}{pq-2}\left(\frac{v_{pq,n}}{v_{1,n}}\right)^{pq/2} \frac{1}{v_{pq,n}}\right\} \\ &\quad \leqslant \mathrm{DMSE}_n \leqslant c\left\{2 - \frac{c}{pq-2}\left(\frac{v_{1,n}}{v_{pq,n}}\right)^{pq/2} \frac{1}{v_{1,n}}\right\},\end{aligned} \tag{6}$$

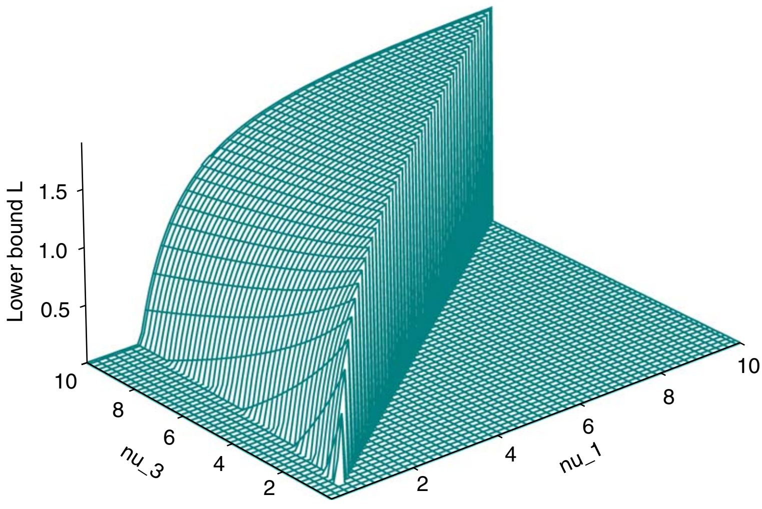

Fig. 1. The lower bound L of (8) with $pq = 3$ and $c = 1$ (truncated by $L = 0$).

which implies that $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ *improves* $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ *if the left-hand side of* (6) *is positive.*

(ii)

$$c\left\{2 - \frac{c}{pq-2}\left(\frac{v_{pq}}{v_1}\right)^{pq/2}\frac{1}{v_{pq}}\right\} \leqslant \lim_{n\to\infty} \mathrm{DMSE}_n \leqslant c\left\{2 - \frac{c}{pq-2}\left(\frac{v_1}{v_{pq}}\right)^{pq/2}\frac{1}{v_1}\right\}, \tag{7}$$

which implies that $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ *improves* $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ *asymptotically if the left-hand side of* (7) *is positive.*

The proofs of the theorems are given in Section 3.

Let us examine the lower and upper bound for $\lim_{n\to\infty} \mathrm{DMSE}_n$ numerically.

Let

$$L \equiv c\left\{2 - \frac{c}{pq-2}\left(\frac{v_{pq}}{v_1}\right)^{pq/2}\frac{1}{v_{pq}}\right\}, \tag{8}$$

$$U \equiv c\left\{2 - \frac{c}{pq-2}\left(\frac{v_1}{v_{pq}}\right)^{pq/2}\frac{1}{v_1}\right\}. \tag{9}$$

In Figs. 1 and 2, we plotted $L = 2 - (v_3/v_1)^{3/2}/v_3$ and $U = 2 - (v_1/v_3)^{3/2}/v_1$, respectively (with $c = 1$ and $pq = 3$). From these figures, we observe that $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ improves $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ if $v_1 \approx v_3$ (near), and v_1 and v_3 are not close to 0. If $v_1 \approx v_3$, and v_1 and $v_3 \nearrow \infty$, then the improvement becomes larger.

Example 1. Suppose that $p = 1$ so $y(t) = \mathbf{B}'\mathbf{x}(t) + \varepsilon(t)$, where $y(t)$ and $\varepsilon(t)$ are scalars. Let the $x_i(t)$'s be divided into s sets, each set corresponding to a frequency λ_l, $l = 1, \ldots, s$, with

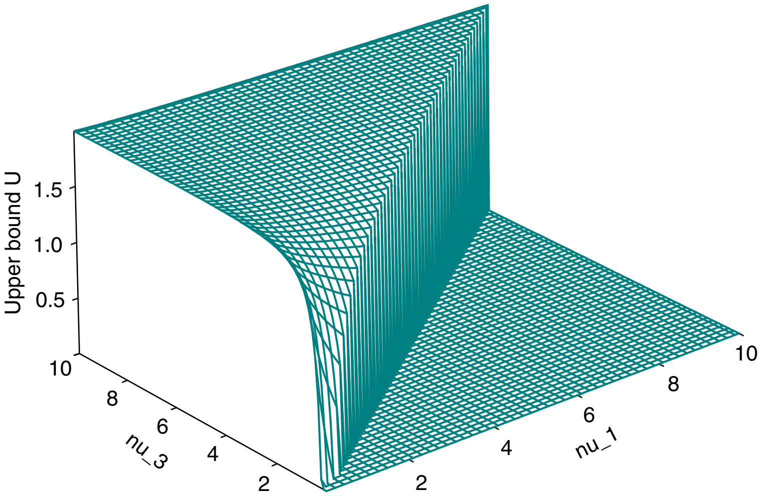

Fig. 2. The upper bound U of (9) with $pq = 3$ and $c = 1$ (truncated by $U = 0$).

$0 \leqslant \lambda_1 < \cdots < \lambda_s \leqslant \pi$ and q_l $x_i(t)$'s corresponding to λ_l. In the set with index l let the $x_i(t)$'s be $t^{k-1} \cos \lambda_l t$, $k = 1, \ldots, q_l$.

Then $\rho_{ij}(h) = 0$ if i and j are in different sets. For i and j in the set with index l, $l = 1, \ldots, s$, $\rho_{ij}(h)$ constitute the $q_l \times q_l$ matrix

$$\mathbf{R}_l \cos \lambda_l h,$$

where the u, vth element of the $q_l \times q_l$ matrix $\mathbf{R}_l$ is

$$\frac{\sqrt{(2u-1)(2v-1)}}{u+v-1} \tag{10}$$

(see Anderson [1, p. 582]). The $\mathbf{M}(\lambda)$ has only jumps. The jump at $\lambda = \pm\lambda_l$ has $\mathbf{R}_l$ as lth diagonal submatrix and 0's elsewhere.

Let us consider two cases.

(i) Suppose that $\{\varepsilon(t)\}$ is generated by MA(1) process such that

$$\varepsilon(t) = u(t) - \theta u(t-1), \quad 0 < \theta < 1,$$

where $u(t)$ is a sequence of i.i.d. random variables distributed as $N(0, \sigma^2)$. Then $\{\varepsilon(t)\}$ has the spectral density

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi} |1 - \theta e^{i\lambda}|^2. \tag{11}$$

In this case, the matrix $\mathbf{C}$ has s diagonal submatrices and 0's elsewhere, and the lth diagonal has elements

$$\sigma^2 |1 - \theta e^{i\lambda_l}|^2 \frac{\sqrt{(2u-1)(2v-1)}}{u+v-1}, \quad u, v = 1, \ldots, q_l.$$

Let $\eta_{1,q_l}, \ldots, \eta_{q_l,q_l}$ $(\eta_{1,q_l} < \cdots < \eta_{q_l,q_l})$ be the eigenvalues of $q_l \times q_l$ matrix $\mathbf{R}_l$. Then, we have

$$v_1 = \min_{1 \leqslant l \leqslant s} \sigma^2 |1-\theta e^{i\lambda_l}|^2 \eta_{q_l,q_l}^{-1}, \quad v_q = \max_{1 \leqslant l \leqslant s} \sigma^2 |1-\theta e^{i\lambda_l}|^2 \eta_{1,q_l}^{-1}.$$

If $q_l \equiv q_0$, $l = 1, \ldots, s$, then

$$v_1 = \sigma^2 |1-\theta e^{i\lambda_1}|^2 \eta_{q_0,q_0}^{-1}, \quad v_q = \sigma^2 |1-\theta e^{i\lambda_s}|^2 \eta_{1,q_0}^{-1},$$

and

$$L = c\left[2 - \frac{c}{q-2}\left\{\frac{|1-\theta e^{i\lambda_s}|^2 \eta_{q_0,q_0}}{|1-\theta e^{i\lambda_1}|^2 \eta_{1,q_0}}\right\}^{q/2} \frac{\eta_{1,q_0}}{\sigma^2 |1-\theta e^{i\lambda_s}|^2}\right], \tag{12}$$

$$U = c\left[2 - \frac{c}{q-2}\left\{\frac{|1-\theta e^{i\lambda_1}|^2 \eta_{1,q_0}}{|1-\theta e^{i\lambda_s}|^2 \eta_{q_0,q_0}}\right\}^{q/2} \frac{\eta_{q_0,q_0}}{\sigma^2 |1-\theta e^{i\lambda_1}|^2}\right]. \tag{13}$$

We evaluated L of (12) in the case with $\theta = 0.1$, $q = 6$, $q_0 = 2$, $c = 0.01$ and $\sigma = 1$ and the case with $\theta = 0.9$, $q = 6$, $q_0 = 2$, $c = 0.01$ and $\sigma = 1$, respectively. Then we observed that $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ improves $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ if $\lambda_1 \approx \lambda_3$ (near), and λ_1 and λ_3 are close to π. If $\lambda_1 \approx \lambda_3$, λ_1 and $\lambda_3 \nearrow \pi$ and $\theta \searrow 0$, then the improvement becomes larger.

(ii) Suppose that $\{\varepsilon(t)\}$ is generated by AR(1) process such that

$$\varepsilon(t) - \theta\varepsilon(t-1) = u(t), \quad 0 < \theta < 1,$$

so that it has the spectral density

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi}|1-\theta e^{i\lambda}|^{-2}. \tag{14}$$

In this case, similarly as the above we have

$$v_1 = \min_{1 \leqslant l \leqslant s} \sigma^2 |1-\theta e^{i\lambda_l}|^{-2} \eta_{q_l,q_l}^{-1}, \quad v_q = \max_{1 \leqslant l \leqslant s} \sigma^2 |1-\theta e^{i\lambda_l}|^{-2} \eta_{1,q_l}^{-1}.$$

If $q_l \equiv q_0$, $l = 1, \ldots, s$, then

$$v_1 = \sigma^2 |1-\theta e^{i\lambda_s}|^{-2} \eta_{q_0,q_0}^{-1}, \quad v_q = \sigma^2 |1-\theta e^{i\lambda_1}|^{-2} \eta_{1,q_0}^{-1},$$

and

$$L = c\left[2 - \frac{c}{q-2}\left\{\frac{|1-\theta e^{i\lambda_s}|^2 \eta_{q_0,q_0}}{|1-\theta e^{i\lambda_1}|^2 \eta_{1,q_0}}\right\}^{q/2} \sigma^{-2} |1-\theta e^{i\lambda_1}|^2 \eta_{1,q_0}\right], \tag{15}$$

$$U = c\left[2 - \frac{c}{q-2}\left\{\frac{|1-\theta e^{i\lambda_1}|^2 \eta_{1,q_0}}{|1-\theta e^{i\lambda_s}|^2 \eta_{q_0,q_0}}\right\}^{q/2} \sigma^{-2} |1-\theta e^{i\lambda_s}|^2 \eta_{q_0,q_0}\right]. \tag{16}$$

We evaluated L of (15) in the case with $\theta = 0.1$, $q = 6$, $q_0 = 2$, $c = 0.01$ and $\sigma = 1$ and the case with $\theta = 0.9$, $q = 6$, $q_0 = 2$, $c = 0.01$ and $\sigma = 1$, respectively. Then we observed that $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ improves $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ if $\lambda_1 \approx \lambda_3$ (near), and λ_1 and λ_3 are close to 0. If $\lambda_1 \approx \lambda_3$, λ_1 and $\lambda_3 \searrow 0$ and $\theta \searrow 0$, then the improvement becomes larger.

Next, we discuss the case of $\boldsymbol{\beta} - \mathbf{b} \neq \mathbf{0}$. Let $d^2(n) \equiv \sum_{i=1}^{q} d_i^2(n)$. Note that $\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b})\|^2 = O(d^2(n))$.

Theorem 2. *In the case of $\boldsymbol{\beta} - \mathbf{b} \neq \mathbf{0}$, suppose that Assumptions 1–5 hold and that $pq \geqslant 3$. Then,*

(i)

$$\mathrm{DMSE}_n = \frac{2c}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b})\|^2}\{\Delta_n(c, \boldsymbol{\beta} - \mathbf{b}) + o(1)\}, \tag{17}$$

where

$$\Delta_n(c, \boldsymbol{\beta} - \mathbf{b}) = \mathrm{tr}\mathbf{C}_n - \frac{2(\boldsymbol{\beta} - \mathbf{b})'(\mathbf{I}_p \otimes \mathbf{D}_n)\mathbf{C}_n(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b})}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b})\|^2} - \frac{c}{2}. \tag{18}$$

Here, we also have the inequality

$$\Delta_n(c, \boldsymbol{\beta} - \mathbf{b}) \geqslant \sum_{j=1}^{pq-1} v_{j,n} - v_{pq,n} - \frac{c}{2}, \tag{19}$$

which implies that $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ improves $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ up to $1/d^2(n)$-order if the right-hand side of (19) *is positive.*

(ii) *Taking the limit of* (19), *we have*

$$\lim_{n\to\infty} \Delta_n(c, \boldsymbol{\beta} - \mathbf{b}) \geqslant \sum_{j=1}^{pq-1} v_j - v_{pq} - \frac{c}{2}, \tag{20}$$

which implies that $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ improves $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ asymptotically up to $1/d^2(n)$-order if the right-hand side of (20) *is positive.*

To examine the sufficient condition for the asymptotic improvement of $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ numerically, let

$$\Delta_0 \equiv \sum_{j=1}^{pq-1} v_j - v_{pq} - \frac{c}{2}. \tag{21}$$

We evaluated Δ_0 of (21) in the case with $q = 3$, $c = 1$ and $v_2 = 0.9v_1 + 0.1v_3$ and the case with $q = 3$, $c = 1$ and $v_2 = 0.1v_1 + 0.9v_3$, respectively. Then we observed that $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ improves $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ if $v_1 \approx v_3$, and v_1 is not close to 0. If $v_1 \approx v_3$, $v_2 \approx v_3$ and $v_1, v_2, v_3 \nearrow \infty$, then the improvement becomes larger.

Example 2. Let $p = 1$ and $x_i(t) = \cos \lambda_i t$, $i = 1, \ldots, q$ with $0 \leqslant \lambda_1 < \cdots < \lambda_q \leqslant \pi$ (Time series regression model with cyclic trend). Then $d_i^2(n) = O(n)$. We find that $\mathbf{R}(h) = \mathrm{diag}(\cos \lambda_1 h, \ldots, \cos \lambda_q h)$, because

$$\rho_{ij}(h) = \begin{cases} \cos \lambda_i h & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Therefore the $\mathbf{M}(\lambda)$ has jumps at $\lambda = \pm\lambda_i$, $i = 1, \ldots, q$. The jump at $\lambda = \pm\lambda_i$ has $1/2$ at the ith diagonal and 0's elsewhere.

Suppose that $\{\varepsilon(t)\}$ is a Gaussian stationary process with spectral density $f(\lambda)$ satisfying Assumption 5. Then it follows that $\mathbf{C} = \mathrm{diag}(2\pi f(\lambda_1), \ldots, 2\pi f(\lambda_q))$. Let $f(\lambda_{(j)})$ be the jth largest value of $f(\lambda_1), \ldots, f(\lambda_q)$. Then

$$v_j = 2\pi f(\lambda_{(j)}), \quad j = 1, \ldots, q.$$

Thus we have

$$\Delta_0 = 2\pi \left\{ \sum_{j=1}^{q-1} f(\lambda_{(j)}) - f(\lambda_{(q)}) \right\} - \frac{c}{2}. \tag{22}$$

Thus, if all the values of $f(\lambda_{(j)})$'s are near and large, then Δ_0 becomes large, hence, in such a case the JS improves the LS greatly. Because our model is very fundamental in econometrics etc., the result on (22) seems important.

Now, we give numerical comparisons between $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ in the models of Example 2. For this, we divide the MSE as

$$\begin{aligned} \mathrm{MSE}_n(\widehat{\boldsymbol{\beta}}) &= E\left[\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2\right] \\ &= E\left[\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}} - E(\widehat{\boldsymbol{\beta}}))\|^2\right] + \|(\mathbf{I}_p \otimes \mathbf{D}_n)(E(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta})\|^2. \end{aligned} \tag{23}$$

The first and second terms of (23) correspond to the variance and squared bias of the estimator, respectively.

In Table 1, we calculate the variances, squared biases and MSE of $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ by Monte Carlo simulations, using the following approximations

$$(\text{variance}) \approx \frac{1}{1000-1} \sum_{j=1}^{1000} \left\| (\mathbf{I}_p \otimes \mathbf{D}_n) \left(\widehat{\boldsymbol{\beta}}_j - \frac{1}{1000} \sum_{k=1}^{1000} \widehat{\boldsymbol{\beta}}_k \right) \right\|^2,$$

$$(\text{squared bias}) \approx \left\| (\mathbf{I}_p \otimes \mathbf{D}_n) \left(\frac{1}{1000} \sum_{k=1}^{1000} \widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta} \right) \right\|^2.$$

Here, $\widehat{\boldsymbol{\beta}}_k$ is the estimate in the kth simulation. We select $n = 100$, $q = 10$ and $\lambda_j = 10^{-1}\pi(j-1)$. These tables show that in each case the bias of $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ is larger than that of $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$. However, because of having much smaller variance, $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ improves $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ with respect to the MSE. The improvement is larger in the case of $\boldsymbol{\beta} - \mathbf{b} = (0.1, \ldots, 0.1)'$ than in the case of $\boldsymbol{\beta} - \mathbf{b} = (1, \ldots, 1)'$.

As a conclusion, it seems very important to consider shrinkage estimators in the estimation of time series regression models.

3. Proofs

In this section, we provide the proofs of the theorems.

Proof of Theorem 1. With $\boldsymbol{\beta} - \mathbf{b} = \mathbf{0}$, we have from (5)

$$\mathrm{DMSE}_n = -c^2 E\left[\frac{1}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b})\|^2}\right] + 2c. \tag{24}$$

Table 1
Numerical comparisons between $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$ of Example 2

	Variance	Squared bias	MSE
(i) $\boldsymbol{\beta} - \mathbf{b} = (0.1, \ldots, 0.1)'$			
(a) MA(1) and $\theta = 0.1$			
$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	9.682093	0.004096453	9.686190
$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$	3.197042	1.728652914	4.925695
(b) MA(1) and $\theta = 0.9$			
$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	15.999432	0.01550308	16.014935
$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$	7.421518	0.91819993	8.339718
(c) AR(1) and $\theta = 0.1$			
$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	10.229084	0.005035216	10.234119
$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$	3.501476	1.600281740	5.101758
(d) AR(1) and $\theta = 0.9$			
$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	107.61319	0.009306226	107.62250
$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$	94.61641	0.160305897	94.77672
(ii) $\boldsymbol{\beta} - \mathbf{b} = (1, \ldots, 1)'$			
(a) MA(1) and $\theta = 0.1$			
$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	9.771357	0.01172009	9.783077
$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$	9.547039	0.11426271	9.661302
(b) MA(1) and $\theta = 0.9$			
$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	16.28119	0.004318206	16.28551
$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$	15.90653	0.105503775	16.01204
(c) AR(1) and $\theta = 0.1$			
$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	10.043619	0.01430085	10.057920
$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$	9.813915	0.16402634	9.977942
(d) AR(1) and $\theta = 0.9$			
$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	108.7131	0.1452452	108.8583
$\widehat{\boldsymbol{\beta}}_{\mathrm{JS}}$	106.7690	0.3445339	107.1135

From (3), it is seen that

$$\mathbf{P}'(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b}) \sim N(\mathbf{0}, \mathbf{V}_n),$$

where $\mathbf{V}_n = \mathrm{diag}(v_{1,n}, \ldots, v_{pq,n})$ and $\mathbf{P}$ is a $pq \times pq$ orthogonal matrix which diagonalizes $\mathbf{C}_n$. Evaluating

$$E\left[\frac{1}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} - \mathbf{b})\|^2}\right] \tag{25}$$

similarly as in Taniguchi and Hirukawa [9], we have

$$\mathrm{DMSE}_n \geqslant c\left\{2 - \frac{c}{pq-2}\left(\frac{v_{pq,n}}{v_{1,n}}\right)^{pq/2}\frac{1}{v_{pq,n}}\right\}. \tag{26}$$

and

$$\mathrm{DMSE}_n \leqslant c\left\{2-\frac{c}{pq-2}\left(\frac{v_{1,n}}{v_{pq,n}}\right)^{pq/2}\frac{1}{v_{1,n}}\right\}. \tag{27}$$

The statement (i) follows from (26) and (27). Taking $\lim_{n\to\infty}$ in (i) we observe (ii). □

Proof of Theorem 2. We can write (5) as

$$\begin{aligned}\mathrm{DMSE}_n &= -c^2 E\left[\frac{1}{\|(\mathbf{I}_p\otimes\mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}-\mathbf{b})\|^2}\right]\\ &\quad +2cE\left[\left\langle(\mathbf{I}_p\otimes\mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}-\boldsymbol{\beta}),\ \frac{(\mathbf{I}_p\otimes\mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}-\mathbf{b})}{\|(\mathbf{I}_p\otimes\mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}-\mathbf{b})\|^2}\right\rangle\right]\\ &= -c^2E_1+2cE_2,\quad \text{(say)}.\end{aligned} \tag{28}$$

We evaluate E_1 and E_2. Let $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} \equiv (\mathbf{I}_p\otimes\mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}-\mathbf{b})$ and $\boldsymbol{\beta}^{(n)} \equiv (\mathbf{I}_p\otimes\mathbf{D}_n)(\boldsymbol{\beta}-\mathbf{b})$.

First, we show that

$$\frac{\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}-\boldsymbol{\beta}^{(n)}\|}{\|\boldsymbol{\beta}^{(n)}\|} = \frac{\|(\mathbf{I}_p\otimes\mathbf{D}_n)(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}-\boldsymbol{\beta})\|}{\|(\mathbf{I}_p\otimes\mathbf{D}_n)(\boldsymbol{\beta}-\mathbf{b})\|} = o(1),\quad \text{a.s.} \tag{29}$$

To avoid unnecessary complexity, without loss of generality, we assume that $p=1$ and obtain

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}-\boldsymbol{\beta}^{(n)} = \mathbf{D}_n(\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}-\boldsymbol{\beta}) = \mathbf{R}_n(0)^{-1}\mathbf{D}_n^{-1}\mathbf{X}'\boldsymbol{\varepsilon},\quad \text{(say)}.$$

Since $\mathbf{R}_n(0)^{-1} = O(1)$ from Assumption 3, we evaluate

$$\mathbf{D}_n^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \left(\frac{1}{d_1(n)}\sum_{t=1}^n x_1(t)\varepsilon_1(t),\ldots,\frac{1}{d_q(n)}\sum_{t=1}^n x_q(t)\varepsilon_1(t)\right)'.$$

By a slight modification of Theorem 4.5.1 of Brillinger [3] or Theorem 3.2 of He [6], there exists a constant $C>0$ such that

$$\limsup_{n\to\infty}\frac{1}{\sqrt{n^{2\kappa_i+1}\log n}}\left|\sum_{t=1}^n t^{\kappa_i}\varepsilon_1(t)\right| \leqslant C,\quad \text{a.s.},$$

which, together with Assumption 2, shows that for each i and for some constant $C'>0$,

$$\left|\sum_{t=1}^n x_i(t)\varepsilon_1(t)\right| \leqslant C'\left|\sum_{t=1}^n t^{\kappa_i}\varepsilon_1(t)\right| = O(\sqrt{n^{2\kappa_i+1}\log n}),\quad \text{a.s.} \tag{30}$$

From (30), it is seen that

$$\left| \frac{1}{d_i(n)} \sum_{t=1}^{n} x_i(t)\varepsilon_1(t) \right| = O\left(\frac{\sqrt{n^{2\kappa_i+1}\log n}}{d_i(n)} \right), \quad \text{a.s.}, \quad i = 1, \ldots, q,$$

which implies (29).

Now, from (29) it is seen that

$$\begin{aligned}
\frac{1}{\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}\|^2} &= \frac{1}{\|\boldsymbol{\beta}^{(n)} + \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}\|^2} \\
&= \frac{1}{\|\boldsymbol{\beta}^{(n)}\|^2 + 2\langle \boldsymbol{\beta}^{(n)}, \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta} \rangle + \|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}\|^2} \\
&= \frac{1}{\|\boldsymbol{\beta}^{(n)}\|^2} \cdot \frac{1}{1 + \frac{2\langle \boldsymbol{\beta}^{(n)}, \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)} \rangle}{\|\boldsymbol{\beta}^{(n)}\|^2} + o(1)} \\
&= \frac{1}{\|\boldsymbol{\beta}^{(n)}\|^2} \cdot \left\{ 1 - \frac{2\langle \boldsymbol{\beta}^{(n)}, \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)} \rangle}{\|\boldsymbol{\beta}^{(n)}\|^2} + o(1) \right\}, \quad \text{a.s.}
\end{aligned}$$

Therefore, using Fatou's lemma, we observe

$$\begin{aligned}
E_1 &= E\left[\frac{1}{\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}\|^2} \right] \\
&= \frac{1}{\|\boldsymbol{\beta}^{(n)}\|^2} \left\{ 1 - \frac{2}{\|\boldsymbol{\beta}^{(n)}\|^2} E\left[\langle \boldsymbol{\beta}^{(n)}, \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)} \rangle \right] + o(1) \right\} \\
&= \frac{1}{\|\boldsymbol{\beta}^{(n)}\|^2} \{1 + o(1)\} \\
&= \frac{1}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b})\|^2} \{1 + o(1)\},
\end{aligned} \tag{31}$$

and

$$\begin{aligned}
E_2 &= E\left[\left\langle \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}, \frac{\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}}{\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}\|^2} \right\rangle \right] \\
&= E\left[\left\langle \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}, \frac{\boldsymbol{\beta}^{(n)}}{\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}\|^2} \right\rangle \right] + E\left[\frac{\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}\|^2}{\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}\|^2} \right] \\
&= E\left[\left\langle \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}, \frac{\boldsymbol{\beta}^{(n)}}{\|\boldsymbol{\beta}^{(n)}\|^2} \left\{ 1 - \frac{2\langle \boldsymbol{\beta}^{(n)}, \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)} \rangle}{\|\boldsymbol{\beta}^{(n)}\|^2} + o(1) \right\} \right\rangle \right] \\
&\quad + E\left[\frac{\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}\|^2}{\|\boldsymbol{\beta}^{(n)}\|^2} \left\{ 1 - \frac{2\langle \boldsymbol{\beta}^{(n)}, \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)} \rangle}{\|\boldsymbol{\beta}^{(n)}\|^2} + o(1) \right\} \right]
\end{aligned}$$

$$= \frac{1}{\|\boldsymbol{\beta}^{(n)}\|^2} \left\{ -2E\left[\frac{\langle \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}, \boldsymbol{\beta}^{(n)} \rangle^2}{\|\boldsymbol{\beta}^{(n)}\|^2} \right] + E\left[\|\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}^{(n)}\|^2 \right] + o(1) \right\}$$

$$= \frac{1}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b})\|^2}$$

$$\times \left\{ -2 \cdot \frac{(\boldsymbol{\beta} - \mathbf{b})'(\mathbf{I}_p \otimes \mathbf{D}_n)\mathbf{C}_n(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b})}{\|(\mathbf{I}_p \otimes \mathbf{D}_n)(\boldsymbol{\beta} - \mathbf{b})\|^2} + \mathrm{tr}\mathbf{C}_n + o(1) \right\}. \tag{32}$$

From (28), (31) and (32), the assertions (17) and (18) follow.

The inequality (19) follows from the fact

$$v_{pq,n} = \max_{\mathbf{v} \in \mathbf{R}^{pq}} \frac{\mathbf{v}'\mathbf{C}_n\mathbf{v}}{\|\mathbf{v}\|^2}. \qquad \square$$

References

[1] T.W. Anderson, The Statistical Analysis of Time Series, Wiley, New York, 1971.
[2] S.F. Arnold, The Theory of Linear Models and Multivariate Analysis, Wiley, New York, 1981.
[3] D.R. Brillinger, Time Series: Data Analysis and Theory, expanded ed., Holden-Day, San Francisco, 1981.
[4] E.J. Hannan, Regression for time series, in: M. Rosenblatt (Ed.), Time Series Analysis, Wiley, New York, 1963, pp. 17–37.
[5] E.J. Hannan, Multiple Time Series, Wiley, New York, 1970.
[6] S. He, Uniform convergency for weighted periodogram of stationary linear random fields, Chinese Ann. Math. Ser. B 16 (1995) 331–340.
[7] W. James, C. Stein, Estimation with quadratic loss, Proceedings of the Fourth Berkely Symposium, vol. 1, 1961, pp. 361–379.
[8] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proceedings of the Third Berkeley Symposium, vol. 1, 1956, pp. 197–206.
[9] M. Taniguchi, J. Hirukawa, The Stein–James estimator for short- and long-memory Gaussian processes, Biometrika 92 (2005) 737–746.